후보영역의 밝기 분산과 얼굴특징의 삼각형 배치구조를 결합한 얼굴의 자동 검출

이칠우[†] · 최정주[††]

요 약

본 논문에서는 복잡한 배경으로부터 자연스런 상태로 촬영된 얼굴들을 검출하는 알고리듬에 대해 기술한다. 이 방법은 얼굴영역이 적당한 크기의 블록내에서는 비교적 비슷한 밝기를 지닌다는 점에 착안하였다. 이 사실을 근거로, 먼저 영상을 계층적으로 블록화하고, 블록들의 밝기가 서로 비슷한 영역을 신속히 얼굴후보로 선택하여, 후보영역 내에서 구체적인 얼굴특징을 찾는 단계적 처리방법을 도입하였다. 후보영역내의 구체적인 특징추출을 위해서 어둡고 좁은 영역을 강조하는 국소 휘도치 변환법을 사용하였으며 최종 판단을 위해서는 얼굴의 각 기관이 갖는 삼각형의 배치구조를 제약으로 사용하였다. 매우 간단한 방법으로 얼굴영역을 처리하였기 때문에 특징점들을 추출할 때 생기는 파라메터 설정문제를 피할 수 있고 그 결과 파라메터 값에 크게 의존하지 않는 안정된 시스템 구현이 가능하다.

Automatic Face Extraction with Unification of Brightness Distribution in Candidate Region and Triangle Structure among Facial Features

Chil-Woo Lee[†] and Jung-ju Choi[††]

ABSTRACT

In this paper, we describe an algorithm which can extract human faces with natural pose from complex backgrounds. This method basically adopts the concept that facial region has the nearly same gray level for all pixels within appropriately scaled blocks. Based on the idea, we develop a hierarchial process that first, a block image data with pyramid structure of input image is generated, and some candidate regions for facial regions in the block image are quickly determined, then finally the detailed facial features; organs are decided. To find the features easily, we introduce a local gray level transform which emphasizes dark and small regions, and estimate the geometrical triangle constraints among the facial features. The merit of our method is that we can be freed from the parameter assignment problem since the algorithm utilize a simple brightness computation, consequently robust systems not being depended on specific parameter values can be easily constructed.

1. 서 론

최근에 들어 컴퓨터를 중심으로 한 정보시스템의 급속한 보급과 더불어 사용자 인터페이스에 대한 중요성이 날로 증가되고 있다. 인터페이스는 사용하기 쉽고 사용자 친화적이어야 하며 이를 위해서는 정보시스템이 사용자에 관한 정보를 능동적으로 취득할 수 있어야만 한다. 이 때문에 시스템과 직접 접촉하지 않고 사용자의 부담을 최대한으로 줄일 수 있는 시각적 인터페이싱 방법(visual interfacing method)

본 연구는 '96년도 학술진흥재단 신진교수연구과제 연구비의 지원을 받아 수행되었음.
† 전남대학교 컴퓨터공학과
†† (주) 어플라이드 엔지니어링

이 많은 기대 속에 연구되어지고 있다. 예를 들어 카메라를 통하여 입력된 영상으로부터 사용자의 신원을 확인하거나, 표정 또는 동작(제스처)을 인식하여 그 결과를 응용시스템에 활용하는 연구들이 진행되고 있다. 이러한 연구를 위해서는 일차적으로 입력영상으로부터 인간의 얼굴을 신속하고 정확히 찾아내는 문제가 선결되어야만 한다.

사람의 얼굴에 관한 정보는 일상생활에서 매우 중요한 것으로 오래 전부터 자동화를 위한 연구가 시작되었다. 그러나 얼굴이란 인간에게 매우 쉽게 인식할 수 있는 정형화된 패턴이면서도, 인종, 표정, 조명조건, 운동, 보는 방향 등에 따라 그 형상이 급변하기 때문에 실용적으로 사용이 가능한 얼굴 인식 시스템을 만드는 것은 매우 어려운 일이다. 본 논문에서는 배경이 복잡한 영상을 입력으로 하여 입력된 영상으로부터 사람의 얼굴을 자동적으로 찾아내는 알고리듬에 대해 기술한다.

2차원 영상을 이용하여 얼굴을 인식하는데는 다양한 방법이 제안되어 있으나 대표적인 것으로는 얼굴영역 화소의 누적분포를 이용하는 법[1], 템플리트 매칭을 이용하는 법[2,3,12], 탄성 템플리트(Elastic Template) 이용하는 법[4,12], 주성분 분석법[5,16], 기하 모우멘트를 이용하는 법[6,17], 신경망을 이용하는 법[8,9,7,13], 국소특징을 이용하는 법[10] 등을 들 수 있다.

화소값의 누적분포를 이용하는 법[1]은 초기 얼굴연구의 대표적인 방법으로 수직 또는 수평방향으로 화소의 계조값(gray value) 또는 에지 정보의 누적분포를 계산하고 누적분포의 극대(소) 위치와 얼굴 각 기관의 대칭성을 고려하여 휴리스틱(heuristic)한 방법으로 각 기관(눈, 코, 입 등)의 위치를 계산하는 방법이다. 이 방법은 영상의 농담정보를 그대로 이용하기 때문에 비교적 안정된 결과를 얻을 수 있지만 얼굴의 방향과 조명조건, 배경 등에 영향을 받기 쉽다.

템플리트 매칭을 이용하는 법[2,3,12]은 얼굴의 전체 또는 일부분을 템플리트로 미리 등록하고 입력영상과 템플리트 매칭을 통하여 얼굴을 찾거나 특정 개인을 식별하는 방법이다. 현실적으로 특징을 이용하는 방법보다 인식결과가 좋고 하드웨어적으로 간단히 처리할 수 있는 장점이 있으나 얼굴의 크기 변화, 조명변화 등에 인식 결과가 민감하게 좌우된다[12].

Yuille 등이 제안한 탄성 템플리트(Elastic Template) 매칭법[4,12]은 각각의 기관별로 그 기관의 형상을 나타내는 몇 개의 수학적 모델을 미리 정하고 모델의 파라메터를 변화시키면서 에너지를 계산하여 에너지가 최소로 되는 파라메터 값으로 형상을 인식하는 방법을 택하고 있다. 그러나 이 방법은 인식결과가 파라메터의 초기치에 영향을 받기 쉽고 각 기관의 위치를 어느 정도 미리 알고 있지 않으면 적용이 불가능하다. 또 사람의 표정, 움직임에 의해 얼굴기관의 형상은 크게 변하므로 얼굴기관의 전체 형상을 하나의 파라메터로 표현하고 있는 한 파라메터의 탐색공간이 매우 넓어질 가능성이 있다.

주성분 분석법[6,16]은 얼굴영상을 하나의 통계적인 데이터를 보고 여러 가지 얼굴 데이터를 수집하여 통계적인 수법에 의해 영상집합의 주성분을 추출하고 입력영상과 주성분영상의 내적이 최대가 되는 방법으로 개인을 식별해낸다. 이 방법은 얼굴영상집합을 크기, 또는 보이는 방향별로 만들면 매우 안정되고 정확한 인식결과를 얻을 수 있으나 템플리트 매칭의 경우와 마찬가지로 얼굴영상의 다양한 변화 즉 조명의 변화와 복잡한 배경 등에 대응하는 일반적인 영상집합을 어떻게 구성할 것인가가 가장 큰 문제로 남는다.

신경망을 이용하는 법[8,9,7,13]은 이 분야 연구의 발전과 함께 얼굴인식에도 폭 넓게 적용되고 있다. 입력 영상으로는 원영상을 그대로 쓰거나 블록처리를 하여 단계적으로 인식하는 방법들이 제안되어 있다. 그러나 이 방법의 문제는 학습 데이터가 지닌 특징에 따라 인식결과가 크게 좌우되어 다양한 변화에 대응하기 어렵다는 점과 항상 얼굴 전체를 인식의 기본 단위로 삼고 있기 때문에 얼굴의 각 기관이 갖는 특징을 인식에 이용할 수 없다는 점이다.

국소특징을 이용하는 법[10]은 얼굴을 눈, 코, 입 등의 얼굴 기관의 조합으로 보고 각 기관이 지닌 공통적인 특징, 즉 특정 형태의 에지나 코너점을 찾아 후보로 등록하고 이를 대상으로 미리 내장된 얼굴 모델과 전체적으로 기하학적 또는 통계적인 제약이 만족될 때 이 코너점이 속한 영역을 얼굴영역으로 판정하는 방법이다. 이 방법은 위에서 설명한 방법들에 대해 buttom-up적인 인식 방법이라고 볼 수 있으나 초기 후보들을 탐색할 때 배경이 복잡하면 수많은 후보가 존재하여 인식이 어렵게되고 또 처음에 누락

된 후보점들이 인식 결과에 결정적인 영향을 미친다는 점이다.

얼굴인식 문제의 어려움은 얼굴변화의 다양성에 있다. 즉 조명의 방향, 두부의 움직임, 카메라 응시방향, 머리카락의 외형, 안면의 표정변화 등에 의해 거의 무한에 가까운 영상으로 변화하기 때문에 이를 몇 가지의 판단 기준에 의해 손쉽게 구별해 낼 수 있는 일반적인 방법은 존재하지 않는다. 이 때문에 얼굴인식의 자동화는 많은 응용분야에서 흥미 있는 연구로 다루어지고 있으나 이러한 다차원 복잡계의 성질을 가진 관계로 아직도 컴퓨터비전 분야의 가장 어려운 문제중의 하나로 남아 있다. 가장 손쉬운 방법은 인식하고자 하는 얼굴 대상을 한정하거나 얼굴이 나타나는 환경에 제약을 가하면 될 것으로 생각되나 살아서 자유롭게 움직이는 인간의 얼굴이라는 점에서 그러한 제약을 적용하면 응용의 범위가 매우 한정된다는 문제점이 있다. 그러나 얼굴은 주변의 배경과는 다른 특이한 패턴이라는 점에서 무언가의 기준에 의해 단일 영역으로 세그멘테이션이 가능하며 얼굴을 복잡한 배경으로부터 정확하게 분리할 수만 있다면 수많은 분야에 응용이 가능하다.

그런 의미에서 근래에 들어 칼라영상을 이용하여 얼굴을 검출하는 방법이 제안되어 그 실효성이 입증되고 있다. 칼라영상은 계조(밝기)정보 이외에도 색상정보를 포함하고 있으므로 얼굴과 같이 고유한 색상을 지닌 영역을 세그멘테이션 하는데 유리하다. 따라서 조명조건과 인종에 따라 매우 다양하게 변하는 얼굴 색상을 기술할 수 있는 모델을 고안하면 유용하게 사용할 수 있다[19]. 그러나 흑백영상에서의 얼굴검출은 칼라 정보를 이용할 수 없는 경우, 또는 흑백정보만을 이용한 간단한 시스템을 제작할 경우에 필요하며, 칼라영상의 단일 색상에 대한 인식 알고리즘으로도 활용이 가능하다는 점에서 흑백영상에 대한 연구의 의미는 크다. 이러한 점들을 고려하여 본 논문에서 흑백영상으로부터 얼굴의 특징을 찾기보다 블록영역의 밝기 분포와 간단한 얼굴의 구조 모델을 결합한 얼굴검출 알고리즘에 대해 기술한다.

얼굴이 특이한 패턴으로 구별되는 원인은 얼굴 내부의 구성요소, 즉 얼굴기관에 있다. 이 기관들이 특이한 형상을 지니면서 모든 사람에 대해 거의 일률적인 배치 구조를 지니고 있어 이 점이 얼굴이라는 물체를 판단하는 기준이 된다. 따라서 모든 기관을 정확히 찾을 수만 있다면 얼굴은 쉽게 배경으로부터 분리될 수 있다. 그러나 얼굴 기관을 나타내는 특징, 즉 에지나 코너점은 전체 영상에 비해 사이즈가 매우 작기 때문에 배경이 복잡해지면 수많은 특징후보들 때문에 탐색공간이 기하급수적으로 증가하여 얼굴 탐색은 복잡한 문제로 발전한다.

이 문제를 해결하는 방법으로는 가능성이 있는 얼굴후보영역을 재빨리 찾아내고 그 후보영역 속에서 얼굴기관의 특징을 찾아 얼굴영역을 확인하면 된다. 즉 복잡한 배경하에서 전체적인 크기와 얼굴구조 때문에 찾기 쉬운 얼굴을 먼저 찾고 그 속에서 세밀한 특징들을 확인하는 것이다. 이는 top-down 적인 방법과 bottom-up 적인 방법을 결합한 것으로 얼굴의 다양성에 대한 약간의 제약을 가하고 얼굴기관이 갖는 기하학적인 배치정보를 이용하면 매우 효율적인 탐색이 가능하게 된다.

본 논문에서는 복잡한 배경을 갖는 한 장의 영상으로부터 시선 축을 중심으로 회전이 되고 조명조건에 따라 밝기가 불균일하게 변하며, 또 촬영거리에 따라 크기가 달라진 얼굴을 신속히 찾아내는 알고리듬에 대해 기술한다. 이는 위에서 설명한 바와 같이 변화의 다양성 때문에 얼굴의 정면이 어느 정도 보인다는 약간의 제한을 두어 문제를 간단히 하였으며 블록단위의 계조(밝기)정보를 이용하여 얼굴의 후보영역을 추출하고 후보영역 속에서 국소영역처리에 의해 얼굴기관을 추출하고 이를 간단한 얼굴모델을 이용하여 검증하였다는 점에서 top-down 적인 방법과 bottom-up 적인 방법을 결합시킨 것이다.

서론 이후의 본 논문의 구성은 다음과 같다. 먼저 2절에서는 다양한 크기의 얼굴 영역을 쉽고 빠르게 검출하기 위한 블록 단위 휘도치 검색 방법의 개요와 적용방법에 대해서 설명하고, 이어 3절에서는 얼굴 영역을 최종적으로 결정하기 위한 과정인, 전처리 과정과 얼굴 특징 영역 추출에 사용된 지역적인 휘도치 변환(Local Gray- level Transformation : LGT)방법의 개념과 사용 방법에 대해서 설명한다. 다음으로 특징점을 검출하는 방법과 이를 이용해서 최종적으로 얼굴 영역을 검출하는 방법을 설명한다. 4절에서는 실험을 통해 제안된 알고리듬의 타당성을 검증하고, 마지막으로 5절에서 제안한 방법에 대한 평가와 앞으로의 연구방향을 중심으로 결론을 맺는다.

2. 계층적 블록처리에 의한 얼굴 후보영역의 결정

2.1 피라미드 구조에 의한 블록영상의 표현

블록이란 일정 크기의 영역에 속한 화소들의 평균값을 구해 그 상위 영상, 즉 보다 함축된 영상을 표현하는데 사용되는 기본 단위를 말한다. 이 표현법은 영상을 구조적으로 표현하는 방법으로 매우 유효할 뿐만 아니라 하드웨어적으로도 실현하기가 매우 쉬워 영상처리에서 널리 사용되고 있다. 얼굴은 각 기관이 지닌 특이한 배치구조 때문에 블록으로 표현하면 데이터 양을 줄이면서도 개인차를 쉽게 알 수 있어 이 방법은 얼굴인식에서도 널리 사용되었다 [14,15].

본 논문에서는 그림1과 같이 피라미드 구조를 갖는 블록을 사용하여 영상데이터를 표현한다. 카메라를 통하여 입력된 영상(640×480화소)을 절반 크기로 축소하여 가장 하위에 놓고 이 영상을 5단계에 걸쳐 식(1)의 계산에 의해 연속적으로 하위 2×2 블록의 평균값을 상위의 한 블록에 기록하여 절반으로 축소된 계층구조를 만든다. 이렇게 하면 맨 상위 영상의 한 블록은 원래영상의 32×32화소 크기에 대응된다.
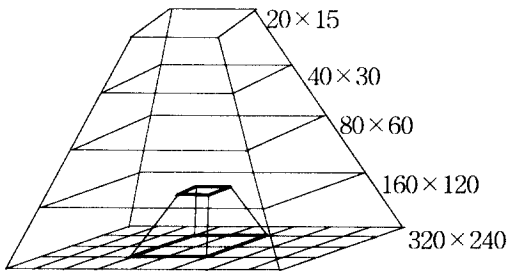
그림 1. 블록영상의 피라미드 구조. 하위 2×2 블록의 평균값을 상위 블록의 값으로 한다. 숫자는 각 계층별 블록영상의 크기를 나타낸다.

$$Block(upper) = \frac{1}{4}\sum_{i=0}^{1}\sum_{j=0}^{1} Block(lower) \quad (1)$$

2.2 저해상도 블록영상에서의 1차 후보영역 추출

그림1과 같은 블록영상을 이용하여 얼굴을 탐색하는 것은 얼굴영상이 비슷한 밝기를 갖는 N×N개의 블록집합으로 표현될 수 있다는 것을 기본 개념으로 한다. 일반적으로 인간이 쉽게 식별할 수 있는 얼굴영상의 최소 크기는 약 32×32화소 정도이나[8], 이 크기의 영역을 배경과 비교하여 볼 때 거의 동일한 휘도치를 갖는 부분 영상임을 알 수 있다. 얼굴은 어떤 크기로 영상에 나타날지 모르기 때문에 그림1의 상위 4개의 블록영상을 대상으로 다음과 같은 계산을 하여 후보영역을 찾는다.

먼저 각 단계의 블록영상에서 전 영역을 스캔하면서 4×4의 영역을 설정한다. 이 4×4영역은 해상도가 두 배인 하위 영상의 블록 평균값으로 만들어진 것이므로 16개 블록의 값이 거의 비슷하게 된다. 어느 경우라도 상위의 영상(하위영상에 비해 해상도가 낮으므로 저해상도 영상이라고도 함)과 하위의 영상(상위영상에 비해 상대적으로 해상도가 높으므로 고해상도 영상이라고 함) 사이에는 이러한 관계가 성립한다.

처리 중인 4×4영역 B_p가 얼굴의 1차 후보영역이 되기 위해서는 식(2)로 정의되는 판단 기준을 만족하여야 한다. 이 식은 영상 일부분의 근사적인 분산을 구하여 그 값이 일정 범위 내에 들어가는 가를 확인하여 영상블록들간의 유사성을 판단한다.

$$B_p = \begin{cases} candidate\ region & : \text{if } Th_1 \le V \le Th_2 \\ non-candidate\ \ region & : otherwise \end{cases} \quad (2)$$
$$where,\quad V = \sum_i\sum_j |B(i,j) - B_{avg}|$$

윗 식에서 Th_1은 블록유사도의 하한 값, Th_2는 상한 값, B_{avg}는 B_p 영역의 평균값을 각각 나타낸다. 또 V는 한 영역의 일반적인 분산값을 의미하지는 않지만 한 블록내의 평균값과 각 화소들의 차의 절대치를 취해 총합을 구함으로써 계산량을 줄이면서 부분영역의 근사적인 분산값을 계산하게 된다. Th_1과 Th_2를 정하는 일반적인 방법은 없으나 실험적으로 이를 구할 수 있다. 표1은 104매의 얼굴영상을 대상으로 하여 각 화소들의 밝기 분산을 계산

표 1. 식(2)의 문턱치 th1, th2를 결정하기 위해 실험에 사용한 얼굴영상을 대상으로 구해 본 얼굴영역 화소의 밝기 분산.

실험에 사용된 얼굴의 수	평균 분산값	분산의 최대값	분산의 최소값	픽셀들의 최대 휘도치분포
104	33	63	8	102 ~ 255

한 결과이다. 이를 근거로 본 논문에서는 Th_1과 Th_2의 값을 최대 밝기 값(255)의 약 10%와 20%로 설정하였으며 이 값들을 조절함에 의해 휘도치가 동일한 배경을 제거하고 1차 후보영역의 개수를 조절할 수 있다.

식(2)를 만족하여 추출된 영역을 1차 후보영역이라고 한다. 이 영역은 최소한으로 얼굴일 가능성을 지닌 영역으로 손쉽게 계산이 가능하다. 그림2는 블록영상의 예를, 그림3은 그림2의 영상을 대상으로 하여 위에서 설명한 방법에 의해 1차 얼굴후보영역을 추출한 결과를 보여준다.

(a) (b) (c)

그림 2. 블록영상의 예. (a)320×240의 원영상. (b)40×30영상 (c)20×15영상

그림 3. 저해상도 영상에서의 1차 얼굴후보영역 탐색결과.

2.3 고해상도 블록영상에서의 2차 후보영역 추출

2.2 절에서는 두 개의 인접하는 영상계층에서 상위 저해상도 블록영상의 밝기값 비교에 의해 개략적으로 얼굴의 1차 후보영역을 추출하는 방법에 대해 설명하였다. 그러나 그림3에서 보는 바와 같이 저해상도에서의 판별은 너무 간단한 기준만을 사용하고 있어 얼굴영역이 아닌 부분도 쉽게 후보영역으로 선택되는 경향이 있다. 물론 식(2)의 Th_1과 Th_2를 세밀히 조절하면 후보수를 줄일 수 있으나 입력영상의 상황변화에 대한 일반적인 설정기준은 없으므로 보다 근본적인 해결책은 해상도가 두 배인 하위의 고해상도 영상에서 1차 후보영역에 대해 보다 구체적인 검사를 실시하는 것이다. 이 절에서는 얼굴패턴에 대한 경험적인 지식을 이용하여 간단히 후보를 줄여가는 방법에 대해 기술한다.

그림2로부터 저해상도영상(그림(c))은 얼굴 또는 물체의 어렴풋한 영역을 나타내지만 하위의 고해상도영상(그림(b))은 얼굴의 특징, 즉 기관의 존재가 어느 정도 나타나 있음을 알 수 있다. 물론 얼굴의 방향, 조명의 방향에 의해 다소의 차이는 발생하지만 정면을 향한 얼굴의 경우 추출된 후보영역이 얼굴부분이라면 눈 영역이 대부분 드러나게 된다. 얼굴패턴에서의 눈에 대한 다음과 같은 성질을 이용하면 1차후보집합으로부터 배경부분을 쉽게 제거할 수 있다. 구체적인 예를 들면 만일 그림2의 20×15 영상(그림2.(c))에서 하나의 4×4블록이 후보영역으로 결정되면, 하위의 고해상도 40×30 영상(그림2.(b))의 해당 영역에서 아래에 기술한 얼굴의 특징을 탐색하게된다.

- 눈은 후보로 선택된 영역에서 가장 낮은 휘도치를 갖는다.
- 두 눈은 후보영역 폭의 약 50~80%의 거리를 두고 떨어져 있다.
- 두 눈 아래쪽 중심부에 있는 2~3개의 블록들은 거의 동일한 휘도치를 갖는다.

이 제약은 얼굴에서 가장 두드러진 특징중의 하나인 눈을 먼저 탐색하여 이를 기준으로 부가적인 처리를 하는데 널리 사용되는 지식이다. 후보영역 내에 가장 휘도치가 낮은 블록을 찾아서 이들간의 거리를 계산하여 두 번째의 조건을 만족하면 최저 휘도치를 갖는 두 블록을 연결한 선에 수직 방향, 즉 두 눈을 연결한 수직방향으로 존재하는 블록의 휘도치를 비교하여 세 번째 조건을 확인한다. 이 때 그림3으로부터 알 수 있는 바와 같이 인접하는 후보영역이 많을 경우 계산이 많아지므로 인접하는 후보영역을 통합하는 과정이 필요하다. 통합하는 방법은 여러 가지를 고려할 수 있으나 본 논문에서는 위에서 사용한 특징을 근거로 하여 매우 간단한 방법을 사용하였다. 즉, 두 눈의 위치가 두 블록 이하의 거리를 지닌 후보집합, 얼굴 영역의 위치가 세 블록 이하의 거리를 지니면서 그 크기의 차이가 두 블록 이하인 후보집합을 구해 후보영역의 왼쪽상단의 좌표를 기준으로 위치

의 평균값을 구하고 하나의 후보로 통합하였다.

그림4는 이상에서 기술한 방법에 의해 2차 얼굴후보영역을 추출한 결과를 보여준다. 이 그림에서 보는 바와 같이 눈에 대한 간단한 제약을 적용함에 의해 현저히 후보영역이 줄어짐을 알 수 있다. 그러나 이 경우도 가해진 제약이 너무 단순하기 때문에 복잡한 배경의 경우 오류를 일으킬 가능성이 매우 많으므로 다음 절에서 기술하는 국소영역처리를 통해 보다 엄밀히 얼굴 여부를 판단할 필요가 있다.

3. 국소 휘도치 변환을 이용한 얼굴특징의 강조

3.1 국소 휘도치 변환(Local Gray-level Transformation : LGT)

2절에서 설명한 방법은 각 블록의 밝기 정보를 이용하여 개략적인 얼굴후보를 선택하였으므로 그 영역 속에 얼굴의 구체적인 특징들이 포함되어 있는가를 확인하는 작업이 필요하다. 1절에서 설명한 바와 같이 얼굴의 주된 특징은 각 기관의 에지와 코너이나 얼굴영역의 밝기가 매끄럽게 변하기 때문에 검출하기는 쉽지 않다. 본 절에서는 후보영역에서의 간단한 계조도 변환을 통해 특징추출을 용이하게 하는 방법에 대해 기술한다.

만일 입력된 영상이 양호한 환경(정면을 향한 얼굴에 카메라 쪽에서 설치된 조명을 사용한 경우)에서 촬영된 것이라면 간단한 시행착오적 방법에 의해 특징영역을 구별해 낼 수 있다. 그러나 일반적으로 실제 세계에서 어떠한 제약도 가하지 않고 획득한 영상은 대개 어둡거나 때로는 너무 밝아서 얼굴의 특징이 제대로 분리되지 않는 경우가 발생한다. 그럼에도 불구하고 얼굴의 눈, 코, 입은 안면 부위에서 상대적으로 어두운 밝기를 가진다. 문제는 다른 영역과 비교해 상대적인 차이가 매우 작아서 다른 영역으로 구별하는데 필요한 문턱치 설정이 매우 어렵다는데 있다. 이를 손쉽게 하는 방법은 국소 휘도치 변환(LGT)을 이용하여 그 차이를 강조하는 것이다.

LGT는 일반적인 얼굴에서 각 기관에 해당하는 영역이 상대적으로 어둡고 좁은 영역, 즉 화소의 수가 적은 영역이라는 사실을 이용한다. 얼굴영역 화소들의 히스토그램을 구하고 임의의 그레이레벨을 g라고 하면 LGT에 의한 변환은 식(3)과 (4)로 표현된다.

$$I(g)_{TMP} = \frac{\alpha}{n_t} \sum_{i=0}^{g} i \times n_i \tag{3}$$

$$I(g)_{LGT} = \begin{cases} I(g)_{TMP} & : \text{if } I(g)_{TMP} \leq 255 \\ 255 & : otherwise \end{cases} \tag{4}$$

여기서, $I(g)_{TMP}$는 변환을 위해 도입한 임시 변수이고, $I(g)_{LGT}$는 LGT에 의해 변환된 얼굴 후보영역 내 그레이레벨 g의 새로운 값을 나타낸다. α는 강조의 정도를 나타내는 상수이며 본 논문에서는 α의 값을 그레이레벨의 최고 값인 256을 사용하였다. 또 n_t는 후보영역 전체의 픽셀수이고, n_i는 i값의 밝기를 갖는 픽셀들의 수이다. 얼굴영역에서 눈, 코, 입과 같은 기관은 상대적으로 좁은 면적을 차지하므로, 위 식은 후보영역 속에서 어둡고 좁은 각 기관부분의 밝기는 어두운 값을 유지하는 반면, 넓고 밝은 안면부분은 상대적으로 급격히 밝은 값으로 변환시킴에 의해 얼굴의 특징영역과 다른 부분의 차별화를 가해 분리가 용이하게 한다.

그림5는 LGT의 변환특성을, 그림6은 실제 후보영역 중 눈이 존재하는 부분을 수평으로 잘라내서 LGT 처리한 후의 결과를 보여주고 있다. LGT에 의해 문턱치의 설정이 용이해짐을 알 수 있다. 또 그림7은 그림3의 한 후보영역에 LGT를 가한 결과를 보여준다.

(a) (b) (c)

그림 4. 고해상도 영상에서의 얼굴후보영역 추출결과. (a)블록영상. (b)인접 블록의 통합 이전의 추출결과. (c)근접 후보영역을 통합하고 영역을 확대시킨 결과
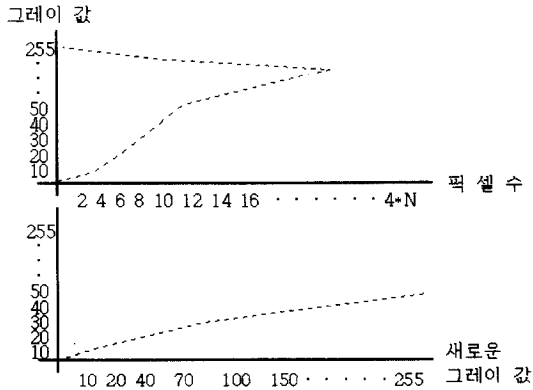
그림 5. 후보영역의 히스토그램과 LGT의 변환 특성
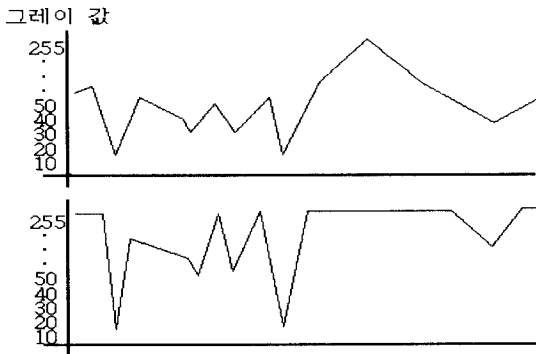
그림 6. 눈이 존재하는 부분의 밝기 변환. LGT 처리전(위) 와 LGT 처리후(아래)

3.2 특징영역의 레이블링

후보영역이 얼굴인가의 여부는 1절에서 설명한 바와 같이 그 영역 내에 얼굴의 특징, 즉 눈, 코, 입이 존재하는가에 달려있다. 이 판단을 위해서는 영역 내의 분리된 부분이 각기 독립된 영역으로 레이블링(labeling)되고 이들간의 관계를 명확히 함이 필요하다. 본 논문에서는 이후에 설명하는 기하 모델을 적용시키기 위해 다음과 같이 레이블링을 실시하였다.

그림3의 맨 하위 레벨 영상의 후보영역에 대해 눈영역을 중심으로 추출된 후보영역보다 2블록정도 넓게 영역을 새로 설정하고 가우시안 필터링을 실시한후 LGT를 가하여 이진영상으로 변환한다. 2진화를 할 때 문턱치를 설정하는 문제가 발생하나 LGT에 의해 설정 범위가 확대된 관계로 시행착오적인 방법을 택해도 좋고 성능이 평가된 일반적인 자동설정 방법을 택해도 무방하다. 본 논문에서는 Otsu가 제안한 방법[18]을 사용하였다.

그림8은 위와 같은 방법으로 이진화된 영상을 보인다. 이 영상에서 일정 크기 이하의 영역을 제거하고 블록화에 의해 이를 넓은 영역으로 통합하고 레이블링하면 그림9와 같은 결과를 얻는다. 이 그림에서 두 눈의 영역이 매우 확실하게 분리됨을 알 수 있다. 이 영상을 근거로 다음 절에서 설명하는 바와 같이 눈과 입이 될 수 없는 영역들은 제거된다.

그림 7. 그림3의 한 후보영역을 LGT처리한 후의 결과

그림 8. LGT 처리후의 영상(그림7)을 Otsu 알고리즘에 의해 이진화한 영상

그림 9. 이진영상을 블록화하여 레이블링한 후의 결과영상

3.3 삼각형 구조 모델을 이용한 얼굴 영역 결정

얼굴의 각 기관부분이 특징으로 추출된 후에는 이 영역들이 배치된 구조가 얼굴을 결정하는 중요한 단서가 된다. 배치구조는 얼굴 전체가 3차원 물체인 관계로 3차원 구조를 가지나 두 눈이 보일 정도로 정면을 향한 얼굴은 2차원으로 간략화할 수 있다[7]. 본 논문에서는 그림10에서 보이는 바와 같은 간단한 삼각형 모델을 얼굴모델로 사용하여 얼굴영역을 결정한다.

그림10에서 ER은 오른쪽 눈, EL은 왼쪽 눈, M은 입을, $\angle ER$, $\angle EL$, $\angle M$은 양 눈과 입을 연결한 선분들의 사이각을 각각 나타낸다. 표2는 이들 각 특징 영역간의 기하학적인 위치관계를 양 눈간의 거리와
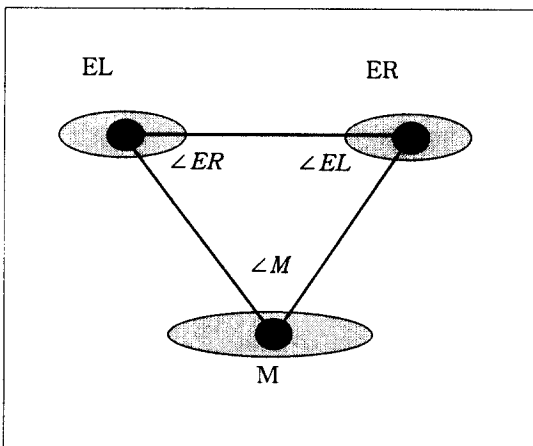
그림 10. 2차원으로 간략화한 얼굴의 기하모델.

표 2. 그림9의 얼굴모델이 갖는 기하제약

특징조건	크기(거리/각도)	허용오차
1) ER_EL	1.0(기준거리)	0.0
2) EL_M	1.23	0.25
3) ER_M	1.23	0.25
4) $\angle ER$	1.0(기준각)	0.1
5) $\angle EL$	1.0	0.1
6) $\angle M$	0.83	0.2

눈영역의 사이각을 1로 하여 나타낸 값이다. 이 값들은 실험에 사용한 얼굴 데이터를 대상으로 화면상에서 실측을 통해 얻었다. 또 허용오차는 얼굴의 회전과 크기 변동에 의해 발생할 수 있는 변화량으로 이 오차를 고려하면 대부분의 얼굴이 만족할 수 있는 기하학적인 제약으로 사용이 가능하다. 얼굴의 후보영역에서 레이블링된 영역이 이 조건을 만족하면 그 후보는 얼굴로 판정된다.

후보영역 내에서 레이블링된 영역들이 위의 조건을 만족시키는가는 다음 순서에 의해 판단한다. 먼저 2.2절의 고해상도 탐색과정에서 얻은 눈 후보영역에 해당하는 특징영역을 선택하여 일정 넓이를 지니고 있는가를 확인하고 중심점을 결정한 후 표2의 특징조건 1)을 계산한다. 두 눈을 연결하는 수직방향으로 레이블링된 특징영역을 선택하여 조건 2) ~ 6)을 만족하면 얼굴의 후보 영역으로 선택한다. 위의 조건중 하나라도 만족하지 않을 경우는 얼굴영역이 아니므로 즉시 후보영역에서 제거한다.

4. 실험 결과

이상에서 설명한 알고리즘의 성능을 확인하기 위해 실험실 내에서 촬영한 60매의 영상을 대상으로 실험을 행하였다. 영상은 실내 전등이 켜 있는 상태에서 얼굴이 임의의 방향과 크기로 포함되도록 촬영하였으며 어떠한 제약도 가하지 않았다. 비디오 카메라로 촬영된 흑백의 640×480 영상을 반으로 축소하여 그림1의 최하위 320×240 영상 데이터를 만든다. 이 영상을 근거로 상위의 블록영상을 만들어 얼굴영역의 추출에 사용한다. 모든 실험은 Sun solaris 2.0

을 운영체제로 하는 Sun SPARC station-20을 사용하였으며 한 장의 영상을 전부 처리하여 인식하기까지는 약 2.3초의 시간이 걸렸다.

실험의 결과에 대한 몇 가지 예를 그림11, 그림12에 보이고 종합적인 결과는 표3에 정리하였다. 여기서 얼굴이 기울어지거나 회전된 경우는 얼굴의 두 눈이 사라지지 않을 정도의 기울기나 회전을 뜻한다. 이런 경우 표3에서 보는 바와 같이 성공률이 다소 떨어지나, 대체적으로 만족할 만한 결과를 얻을 수 있음을 알 수 있다.

(a) (b)

(c) (d)

그림 11. 영상에 하나의 얼굴이 존재할 경우의 실험결과. (a) (b) 얼굴이 똑바로 선 경우의 검출 결과. (c) 얼굴의 가로로 약 20° 회전했을 경우의 얼굴 검출 결과. (d) 얼굴이 세로로 약 30° 기울어졌을 경우의 실험 결과

(a) (b)

그림 12. 영상에 얼굴이 하나 이상 존재할 경우의 실험결과. (a) 얼굴이 세개인 경우의 검출 결과. (b) 두개의 얼굴이 기울어진 경우의 실험 결과

표 3. 얼굴영상 60매를 대상으로 한 인식결과

	실험 영상 수	영상속의 얼굴 수	검출된 얼굴 수	잘못 검출된 얼굴 수
정면을 향한경우	30	60	56	4
기울어지거나 회전된 경우	30	62	50	5

5. 결 론

본 논문에서는 영상의 배경, 얼굴의 크기와 회전, 기울기 등에 제약을 가하지 않고 자연스러운 조건에서 획득한 영상으로부터 서로 다른 기울기와 회전 정도를 갖는 얼굴들을 검출하는 알고리즘을 제안하였다. 제안된 방법은 얼굴영역이 비교적 비슷한 밝기를 지닌다는 점에 착안하여 영상을 계층적으로 블록화하고 블록들의 밝기가 서로 비슷한 영역을 얼굴후보로 선택하여 후보영역 내에서 구체적인 얼굴특징을 찾는 단계적 처리방법을 도입하였다. 구체적인 특징 추출을 위해서 어둡고 좁은 영역을 강조하는 국소 휘도치 변환법을 사용하였으며 최종 판단을 위해서는 얼굴특징이 갖는 간단한 삼각형 모델의 기하학적인 제약을 특징점후보에 적용 판단하였다.

제안된 방법의 특징은 블록의 밝기 정보를 기반으로 하여 매우 간단하게 얼굴을 검출하였기 때문에 에지나 코너와 같은 형상 특징들을 추출할 때 생기는 파라메터 설정문제를 피할 수 있다는 점이다. 일반적으로 영상인식에 있어서 가장 문제가 되는 파라메터 값은 환경변화에 매우 민감하게 작용하기 때문에 가능하면 그 수를 줄이거나 값의 변화에 영향을 적게 받는 알고리즘을 고안하는 것이 무엇보다 중요하다. 또 얼굴의 최종 판단을 위해 사용된 삼각형 모델은 그 구조가 간단하여 에러를 발생하기 쉬우나 후보영역의 추출과정에서 얻어진 결과 영역을 이 모델에 결합함으로써 인식과정에서 상승효과를 가져온다는 점도 중요한 특징으로 작용한다. 본 논문에서 제안하는 방법은 처리의 흐름이 영상의 밝기 변화에 대한 간단한 계산만으로 이루어진 관계로 하드웨어로 구현하기가 쉽고 안정성이 높아 많은 활용이 기대된다.

참 고 문 헌

[1] T. Kanade, “Picture Processing by Computer Complex and Recognition of Human Faces”, Tech.. Rep., Kyoto Univ., Dept. Inform. Sci., 1973

[2] R. J. Baron, “Mechanism of Human Facial recognition”, Int. J. Man Machine Studies, Vol. 15, pp.137-178, 1981

[3] Young Ho Kwon and Niels da Vitoria Lobo, “Face Detection Using Templates,” Proc. of ICPR, pp.764-767, 1994

[4] A. Yuille, P. Hallinan, and D. Cohen, “Feature Extraction from Faces Using Deformable Templates”, Int. J. of Computer Vision, Vol.8, No.2, pp.99-111, 1992

[5] M. Turk and A. Pentland, “Eigenfaces for Recognition”, J. of Cognitive Neurosci., Vol.3, No.1, pp.71-86, 1991

[6] Z. Hong, “Algebraic Feature Extraction of Image for Recognition”, Pattern Recognition, Vol.24, No.3, pp.211-219, 1991

[7] G. Cottrell and M. Fleming, “Face Recognition Using Unsupervised Feature Extraction”, In Proc. Int. Neural Network Conference, 1990

[8] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, “Rotation Invariant Neural Network -Based Face Detection,” Proc. of ICPR, pp.38-44, 1998

[9] Paul Juell and Ron Marsh, “A Hierarchical Neural Network for Human Face Detection.” PR, Vol.29, No.5, pp.781-787, 1996

[10] X. Song, C. Lee, X. Gang, and T. Saburo, “Extracting Facial Features with Partial Feature template”, in Proc. of ACCV'93, pp.751-754, Osaka, Japan, 1993

[11] R. Chellappa. et al, “Human and Machine Recognition of Faces: A Survey”, Proc. of The IEEE, Vol.83, No.5, pp.704-740, 1995

[12] R. Brunelli and T. Poggio, “Face Recognition: Features versus Templates”, IEEE Trans. Pattern Anal. & Machine Intell., Vol. 15, No.10,

1042-1052, 1993

[13] J. M. Vincent, J. B. Waite and D. J. Myers, "Precise Location of Facial Features by a Hierarchical Assembly of Neural Nets", Second Int'l Conf. on Artificial Neural Networks, pp. 69-73, 1991

[14] Guangzheng Yang and Thomas S. Huang. "Human Face Detection in a Scene" Proc. CVPR, pp.453-457, 1993

[15] Guangzheng Tang and Thomas S. Huang, "Human Face Detection in a Complete Background", Pattern Recognition. Vol. 27, No. 1, pp.53-63(1994)

[16] A. Pentland, B. Moghaddam, T. Starner, and M. Turk, "View-Based and Modular Eigenspaces for Face Recognition", in Proc. IEEE Coputer Soc. Conf. on Computer Vision and Patt. Recog. 1994, pp.84-91.

[17] V. Govindaraju, S. N. Srihari, and D. B. Sher, "A computational model for face location", in Proc. 3rd Int. Conf. on Computer Vision, 1990, pp.718-722.

[18] N. Otsu, "A Threshold Selection Method From grey-level Histogram", IEEE Transactions Systems, Man and Cybernetics, Vol.9 No.1, pp.62-66, 1979

[19] H. Wu, Q. Chen, and M. Yachida, "Face Detection From Color Images Using a Fuzzy Pattern Matching Method", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.21, No.6, pp.557-563, 1999

이 칠 우

1986년 중앙대학교 공과대학 전자공학과 졸업
1988년 중앙대학교 대학원 전자공학과 졸업(공학석사)
1992년 동경대학 대학원 전자공학과 졸업(공학박사)
1992년~1995년 이미지 정보과학연구소 수석연구원 겸 오사까대학 기초공학부 협력연구원
1995년 리츠메이칸대학 특별초빙강사
1996년~현재 전남대학교 공과대학 컴퓨터공학과 교수
관심분야 : 컴퓨터비젼, 컴퓨터그래픽스, 멀티미디어 데이터베이스

최 정 주

1997년 전남대학교 공과대학 컴퓨터공학과 졸업
1999년 전남대학교 대학원 컴퓨터공학과 졸업(공학석사)
현재 (주)어플라이드 엔지니어링 근무
관심분야 : 컴퓨터비젼, 휴먼인터페이스